Unsupervised Video Object Segmentation for Deep Reinforcement Learning

Authors: Vik Goel, Jameson Weng, Pascal Poupart
Presenter: Siliang Huang

Outline

- Problem tackled
- Solution proposed
- RL background
- Architecture and methods of proposed solution
- Experiments
- Conclusion and future work

Problem: need good image encoder

- For tasks with image inputs, RL algorithms have great performance on many of them. For example, RL outperforms humans on most Atari games.
- To exploit the success of those RL algorithms, we need to feed them good representations of image/input
- Drawbacks of existing imaging processing techniques or image encoder:
  - Require manual input (such as handcrafting features)
  - Assume object features and relation are directly observable from environment
  - Require domain information, or labeled data
  - Convolutional neural network doesn’t need manual input, but it requires more interactions with the environment to learn what features to extract

Solution

- Motion-Oriented REinforcement Learning (MOREL)
    - A novel image encoder to learn good representation
    - The encoder automatically detects and segments moving objects. Then infer the object motion
    - Fully unsupervised
    - No domain information or manual input required
    - Can combine with any RL algorithm
    - Reduced the amount of interaction
    - The learned representations can help RL to come up with policy based on moving objects
    - More interpretable policy
    - Tested performance on all 59 Atari games available

Only moving objects?

- Assumption: position and velocity of moving objects are important, and should be taken into account by an optimal policy
- Some fixed objects are important too (such as treasure, landmine)
- MOREL combines the moving-object encoder with a standard convolutional neural network to extract complementary features

RL background

- Policy gradient techniques
  - Asynchronous advantage actor critic (A3C)
  - Synchronous variant (A2C)

Pop quiz: What is the difference between them? Which one did we play with in Assignment 2?

RL background

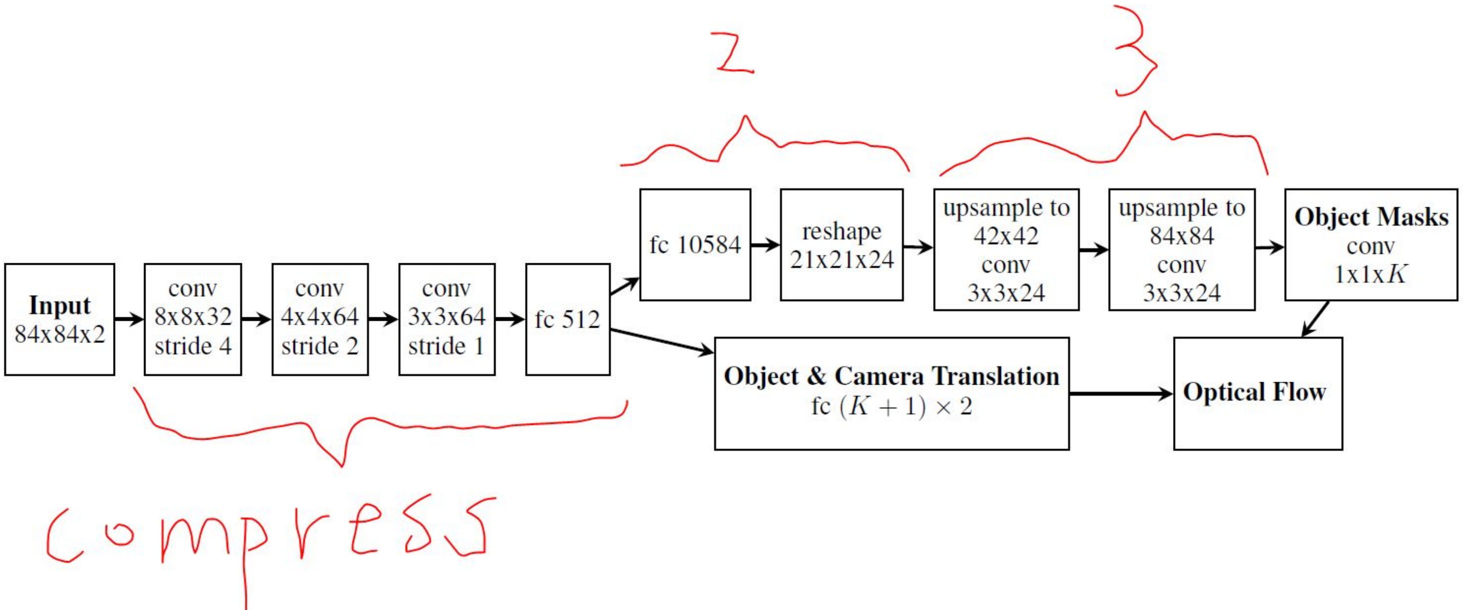
- Policy gradient techniques
  - Asynchronous advantage actor critic (A3C): run multiple copies of same agent in parallel. At update time, pass gradients to a main agent for param updates, then all other agents copy the params of main agent.
  - Synchronous variant (A2C)
  - Problems: gradient might not point to the best direction. Large step size.
- To mitigate those problems
  - Trust region methods
  - Proximal policy optimization (PPO) techniques: clip the policy gradient to prevent overly large changes to the policy.

Overall process of MOREL

- Phase one: the moving object encoder captures structured representation of all moving objects
- Phase two: feed the representation to the RL agent. Continue to optimize the encoder along with optimizing the RL agent.
  - The RL agent will focus on moving objects.
  - The 2nd phase requires less interaction with environment.

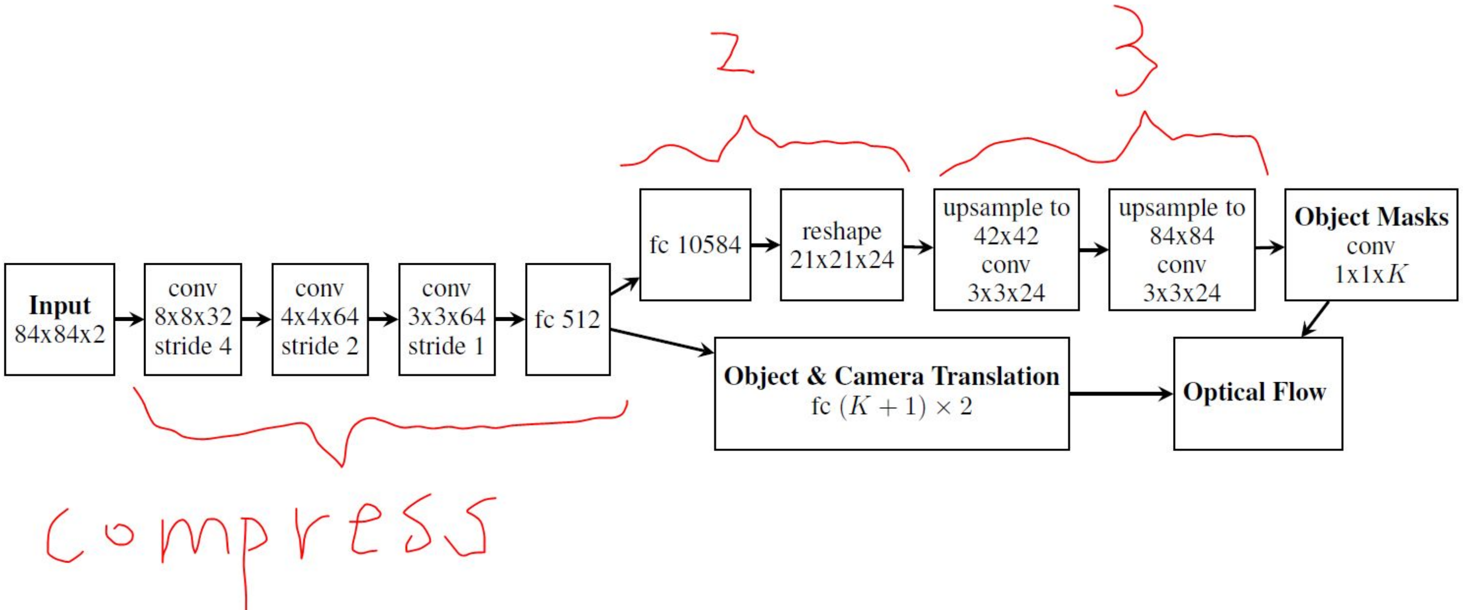
Unsupervised Video Object Segmentation

- This structure is a modified version of Motion Network (SfM-Net)
- Predicts K object segmentation masks
- Each mask has a object translation and a camera translation



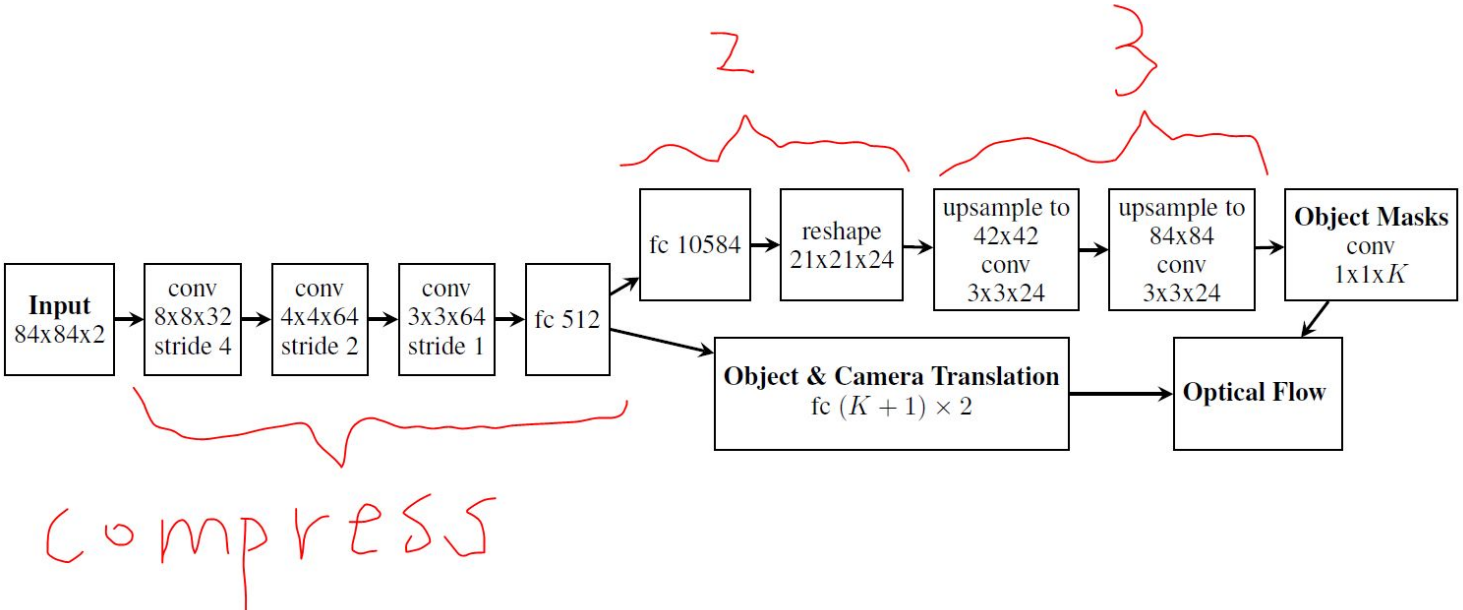
Unsupervised Video Object Segmentation

- Takes 2 frames as input
- Compresses the input images to a 512-dimensional embedding
- 2: reshape activation to a different volume



Unsupervised Video Object Segmentation

- 3: increase size of activations to desired dimensionality for object masks
- A separate flow to compute camera translation
- No skip connection from downsampling path to upsampling path



Object masks

Quality of object masks

- We don’t have ground truth
- We use Reconstruction Loss: estimate the optical flow of the 2nd input image, use that optical flow to wrap the 2nd input image into an estimate of the 1st input image (reconstruction).
- Train the network to minimize the loss between reconstructed estimate and the 1st input image

$$F_{ij} = \sum_{k=1}^{K} (M_{ij}^{(k)} \times t_k) + c$$

Loss function for reconstruction

- We choose structural dissimilarity (DSSIM) loss function, instead of L1.
- The gradient of L1 only depends on immediate neighbouring pixels. Gredient locality problem.
- DSSIM an 11 * 11 filter to ensure gradient at each pixel gets signal from a large number of pixels in its vicinity

Flow Regularization

- Solely minimizing reconstruction loss is not enough. The network can get the correct optical flow while multiple wrong translations cancel out each other.
- One solution: impose L1 regularization on the object masks to encourage sparsity
- Another problem: can obtain correct optical flow with undesirable solution (masks with small values coupled with large object translation)
- Solution: Apply L1 regularization *after* multiplying each mask by its corresponding translation.

$$\mathcal{L}_{reg} = \sum_{k=1}^{K} \|M^{(k)} \times t_k\|_1$$

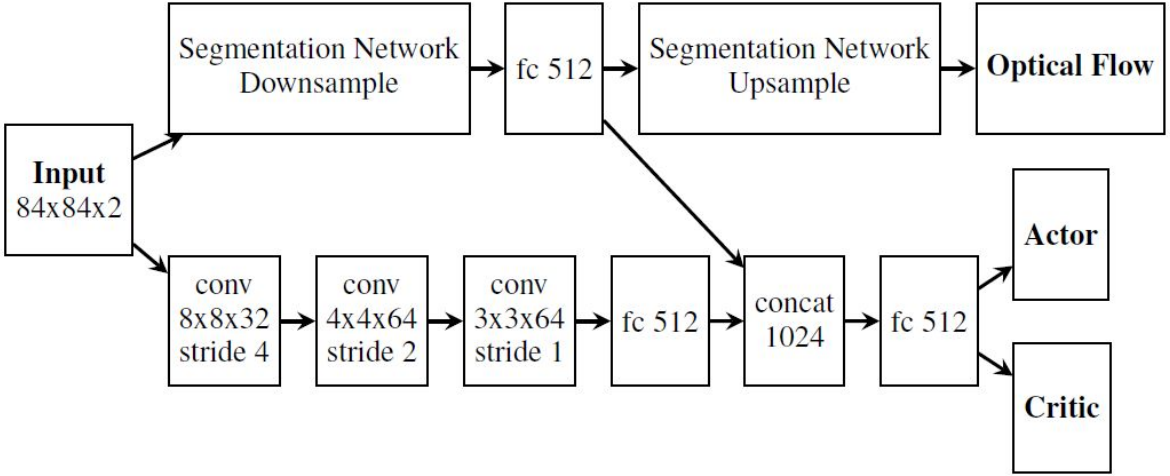
Curriculum

- Minimize segmentation loss with hyperparam lambda.
- Gradually increase lambda from 0 to 1 to make the object mask interpretable without collapsing.

$$\mathcal{L}_{seg} = \mathcal{L}_{reconstruct} + \lambda_{reg}\mathcal{L}_{reg}$$

Phase 2: Transferring for Deep RL

- RL agent needs info about both moving and fixed objects, while the encoder is designed and trained to capture moving objects, not fixed objects.
- Solution: add a downsampling network to capture static objects
- Combine info about moving and static objects.



Joint Optimization

- Minimize segmentation loss along with policy and value function
- Benefits
  - Retaining capability of segmenting objects is useful for visualization
  - Keep improving object segmentation path
  - When game difficulty increases, there will be distribution shift in input. Params in phase one encoder become less meaningful.

Experiments

- To show MOREL can be combined with any RL agent, we combined it with A2C and PPO
- Tested performance on all 59 Atari games available
- Boosted performance of A2C for 26 games; decreased performance on 3 games
- Boosted performance of PPO for 25 games; decreased performance on 9 games

Experiment with encoder

- Finds all moving objects in fully unsupervised manner
- Predicts 20 object segmentation masks (K = 20)
- Displays object masks with the highest confident (highest flow regularization penalty)

Experiment with encoder

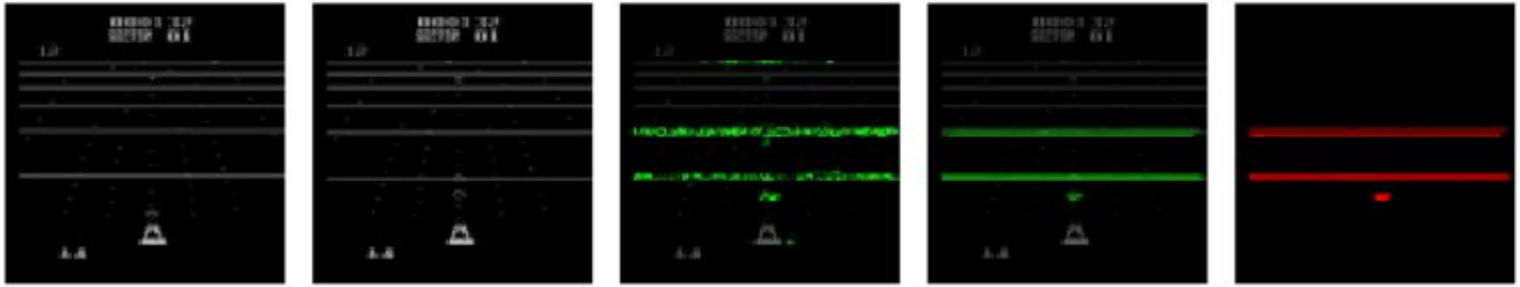
- Deeper green -> more confidence
- Interesting observations: small movement doesn't move pixels in the middle of the object. So the encoder ignores the stationary portions

Experiment with encoder

- Interesting observations: many enemies moves in the same formation. So the encoder puts a mask over all those enemies and treats them as one entity

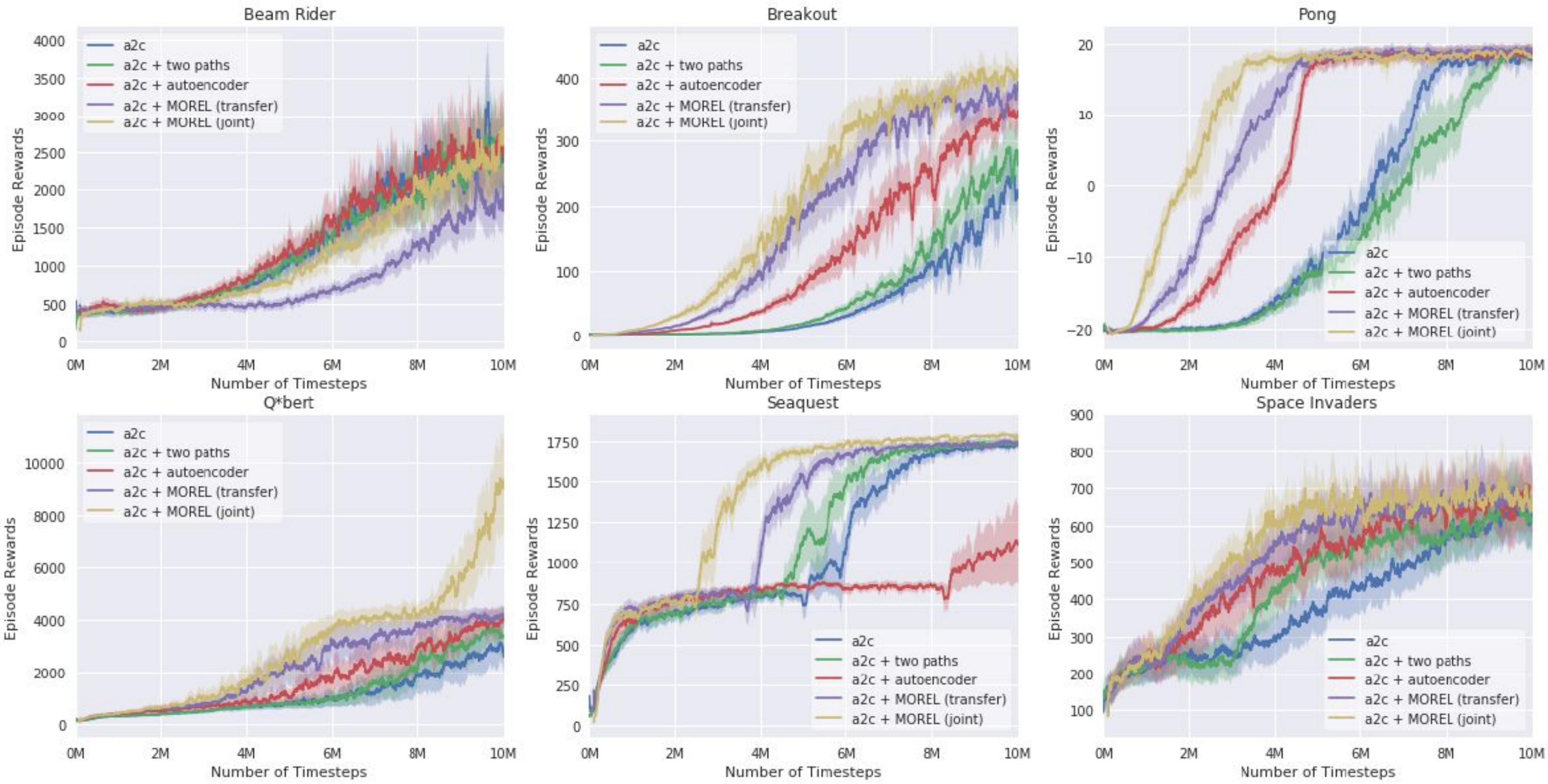
Experiment with encoder

- Interesting observations: For some games, motion is not a helpful cue for understanding the games. Encoder picks up pure visual effects and ignores the smaller enemies. The learned representation is not useful for the RL agent.

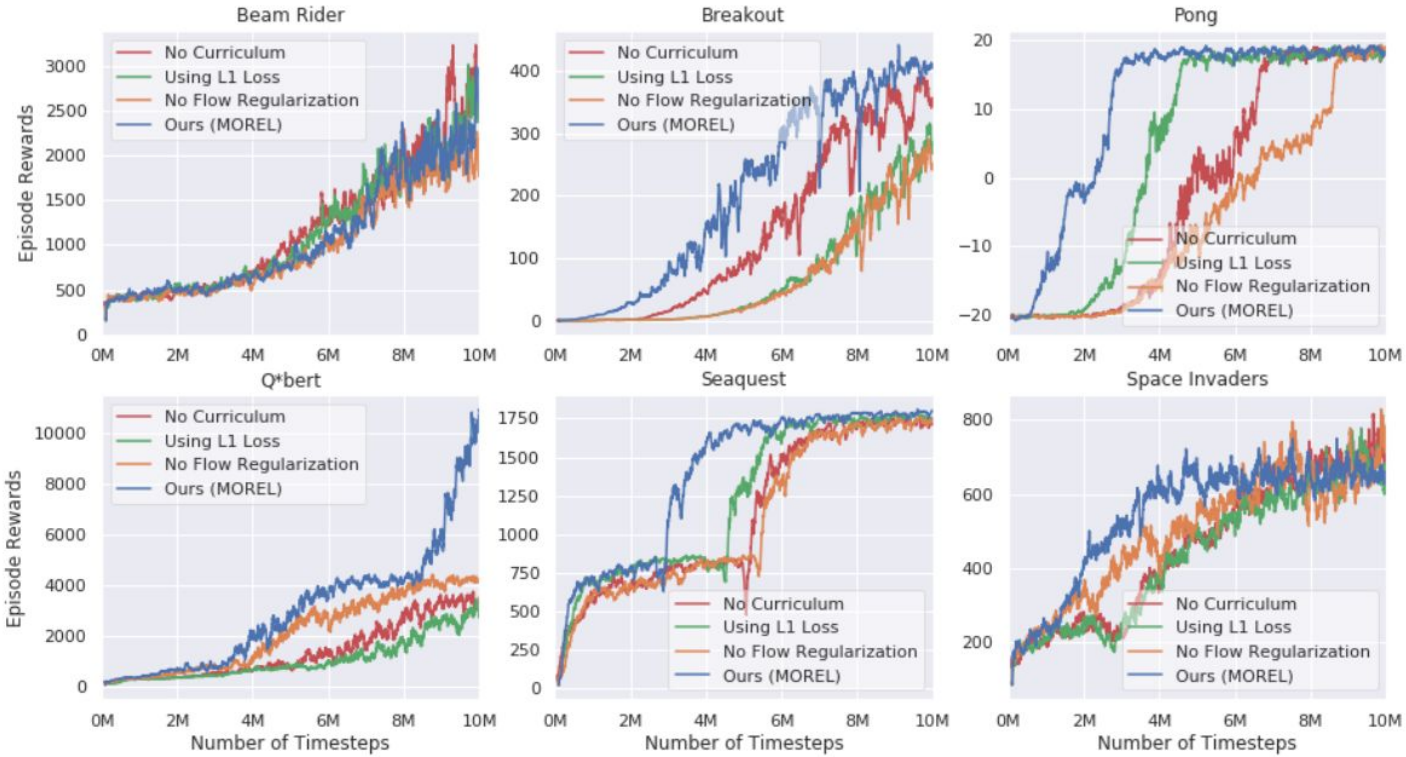


Ablation Study

- Setup:
  - 2 baselines: standard A2C, and A2C with the same architecture as MOREL. Both initialized randomly
  - A2C with autoencoder. Main difference between autoencoder and MOREL is the output. Autoencoder outputs one frame. MOREL outputs K = 20 object masks with object translation and camera motion prediction
  - A2C + MOREL, with and without optimizing jointly
- Results:
  - MOREL didn’t perform worse than baseline in Bean Rider (object mask on visual effect)
  - For Q*bert, optimizing jointly boost the performance significant after reaching 2nd level of the game (never reached during training)


Beam Rider
Breakout
Pong
Q*bert
Seaquest
Space Invaders
Episode Rewards
Number of Timesteps
0M
2M
4M
6M
8M
10M
a2c
a2c + two paths
a2c + autoencoder
a2c + MOREL (transfer)
a2c + MOREL (joint)

Curriculum, flow regularization, DSSIM ablation



Conclusion

- Object segmentation and motion estimation tool
- Advantages:
  - Unsupervised
  - Reduce interaction with environment
  - Can be combined with any RL agent
  - More interpretable policy
- Limitation:
  - Only designed to capture moving object
  - Might ignore small salient moving objects

Future Work

- Extend the encoder framework to fixed objects
- Use attention model to learn salient objects explicitly
- Can combine encoder framework with object-oriented frameworks, physics-based dynamics, model-based reinforcement learning
- Working with 3D environments

Works Cited

- Goel, V., Weng, J., & Poupart, P. (2018). Unsupervised video object segmentation for deep reinforcement learning. In Advances in Neural Information Processing Systems (pp. 5683-5694).

Photo Credit: https://www.pinterest.ca/pin/107523509830651434/